Some Practical Aspects of Free Energy Calculations from Molecular Dynamics Simulation

SRIKANTA SEN, LENNART NILSSON

Center for Structural Biochemistry, Department of Bioscience, Karolinska Institutet, S-14157 Huddinge, Sweden

Received 23 October 1998; accepted 18 January 1999

ABSTRACT: In this work, we address two critical aspects of calculation of the free energy differences in molecular systems from molecular simulations. The first aspect involves checking whether the calculated free energy difference depends significantly on the extent of perturbation used for accomplishment of a given transformation. The second aspect of interest is to verify if the sampling errors in calculating the free energy differences between the wild-type molecule and a mutated one in its free state and in a complex are similar, or not, for a finite-length dynamic simulation. The reliability of the free energy estimates obtained from molecular simulations using thermodynamic cycles depends in part on this fact. For investigating these aspects, we use a self-transformation scheme in which a transformation of a part of a molecular system into itself is considered. We perform MD simulations of DNA fragments in which a part of a specific base is subjected to such a self-transformation. Results indicate that the estimated free energy differences do not depend significantly on the extent of perturbation used to achieve the transformation. Interestingly, the variation in the cumulative free energy difference, ΔA, with the coupling parameter, λ, depends significantly on the extent of perturbation. We examine the physical basis of the observed nature of the variation of the accumulated free energy difference, ΔA, against the λ value in the case of a self-transformation. In a thermodynamic cycle, the sampling errors due to the finite-length simulation for the molecular system are found to be similar to each other for the two perturbations (free and in a complex) justifying the use of such approach in calculating $\Delta\Delta A$ in molecular complexes. J Comput Chem 20: 877–885, 1999

Keywords: free energy perturbation; dynamic simulation; self-transformation; error estimate; dual topology

Correspondence to: L. Nilsson; e-mail: lennart.nilsson@csb.ki.se

Contract/grant sponsor: Swedish Natural Science Research Council

 CCC 0192-8651 / 99 / 080877-09

Introduction

Estimation of free energy differences from molecular simulations has become an important and integrated part of studying the interaction of different molecular systems, particularly in the area of biomolecules.[1–6] In the present work, we examine some critical aspects of the free energy difference calculation from such molecular simulation methods, because it is always important to know how things really appear ''in practice,'' even when their behavior is known ''in principle.'' Here, we have performed free energy calculations by chemical perturbation and molecular dynamics simulation methods using the program package CHARMM,[7] version 25. In the thermodynamic perturbation method, one starts with a hybrid molecular system with a Hamiltonian depending on the molecular coordinates($\mathbf{r}$) and a coupling parameter λ as:

$$H(\mathbf{r}, \lambda) = H_0(\mathbf{r}) + (1 - \lambda) H_A(\mathbf{r}) + \lambda H_B(\mathbf{r}) \quad (1)$$

which is modeled as the sum of $H_0(\mathbf{r})$ and a linear combination of $H_A(\mathbf{r})$ and $H_B(\mathbf{r})$, where $H_0(\mathbf{r})$ is the Hamiltonian of the system excluding the ''reactant'' and the ''product'' parts described by $H_A(\mathbf{r})$ and $H_B(\mathbf{r})$, respectively, evaluated at a given value of the coupling parameter λ through which the transformation takes place. λ can take values in the range (0–1). In the thermodynamic perturbation method, the free energy difference between two states (A and B) can be calculated by using the basic relation[8–11]:

$$\Delta A_{A \to B} = \Sigma[\, A(\lambda_i) - A(\lambda_i + \Delta\lambda)] = -RT \Sigma \ln\langle \exp[\, \beta \Delta H(\lambda_i)]\rangle_{\lambda_i} \quad (2)$$

where:

$$\Delta H(\lambda_i) = H(\lambda_i) - H(\lambda_i + \Delta\lambda) \quad (3)$$

is the difference in the Hamiltonian between the two neighboring states defined by the coupling parameters λ_i and $\lambda + \Delta\lambda_i$; A is Helmholtz's free energy; H is the Hamiltonian, $\beta = (RT)^{-1}$; R is the gas constant; and T is the absolute temperature. The symbol $\langle \cdots\cdots \rangle_{\lambda i}$ denotes a time average of the quantity along the perturbation pathway characterized by the coupling parameter λ_i. Here, the total perturbation is split into a number of smaller ones called windows (between the coupling parameter values λ_i and $\lambda_i + \Delta\lambda$) for which accurate evaluation of the free energy differences by the perturbation method may be possible. The total free energy difference between the two states is then obtained as the sum of the contributions from the individual windows [eq. (2)].

In principle, an infinitely long simulation is necessary for proper sampling all over the phase space available to the molecular system and results obtained in simulations of finite length are generally associated with some error. One way of minimizing error is to perform simulations of sufficient length to ensure the convergence of the calculated free energy value. Recently, Pearlman[12] has shown that, even for a small transition in a small system like ethane, quite long simulation ($\approx$ 700 ps) for each window is required in practice for the convergence of the free energy difference. Thus, estimation of the free energy difference ensuring convergence of values is computationally extremely expensive, particularly for complex biomolecules, and hence, one must remain limited in relatively short, finite-length simulations. Based on these considerations, it is important to check different critical aspects of free energy calculation from molecular simulations when finite length simulations are used. In this work, we address two such critical aspects.

Objectives

The extents of perturbation required to perform the transformation may be widely different starting from a single atom to several atoms. For example, in the perturbative transformation pyrimidine → pyrimidine (cyt → thy or vice versa) in nucleic acids, the perturbation can be done by considering only those atoms that are not present in both the reactant and the product for perturbation, whereas, in the case of purine → pyrimidine transformation, all atoms in the whole base should be changed. So, the first aspect we have to address is to verify whether the calculated free energy difference depends significantly on the extent of perturbation used for accomplishing a given transformation. If this is the case, then the calculated free energy in different cases will be associated with errors of different degrees depending on the extent of perturbations used for the respective transformation, and hence, may need appropriate correction.

The second aspect addressed here relates to the calculation of difference in free energy difference ($\Delta\Delta A$), characterizing the stability difference in a biomolecular system (e.g., DNA–protein complex) due to a mutation by using MD simulation in a thermodynamic cycle as follows:

$$\begin{array}{ccc} & \Delta A_1 & \\ \text{DNA} + \text{protein} & \rightarrow & \text{DNA–protein} \\ \downarrow \Delta A' & & \downarrow \Delta A'' \\ \text{DNA}' + \text{protein} & \rightarrow & \text{DNA}'\text{–protein} \\ & \Delta A_2 & \end{array}$$

where DNA′ is the DNA with the mutation. Because the free energy is a thermodynamic state function, its value does not depend on the actual path of transformation. Thus, in the above thermodynamic cycle one obtains $\Delta\Delta A = \Delta A_1 - \Delta A_2 = \Delta A' - \Delta A''$; that is, $\Delta\Delta A = \Delta A_1 - \Delta A_2$ can also be expressed equivalently as the difference $\Delta A' - \Delta A''$, which is more easily obtained from molecular simulations. However, in such cases, the accuracy of the estimated $\Delta\Delta A$ value depends not only on the quality of the potential energy function used but also on the assumption that the errors introduced due to the simulation protocol used in estimating ΔA, are the same for both the systems (i.e., $\Delta A'$ in the free DNA and $\Delta A''$ in the complex). The validity of this assumption in practice has not been checked explicitly. Because the chemical environments in the immediate neighborhood of the mutation are generally quite different in the free state and in the complex, and also the system sizes can be quite different in the free and complex phases, the possibilities of nonnegligible differences in errors due to finite-length simulation cannot be ruled out. So, in this study, we have also tried to verify this assumption.

Methods

SELF-TRANSFORMATION

To answer the two aforementioned questions, we used a ''self-transformation'' scheme of a molecular system in which a perturbative transition from a molecular system, X, to the same molecular system (X) in the same state is considered. A self-transformation scheme has two unique features: (i) different extents of the molecular system, ranging from a small portion to the whole system, can be used as the perturbing part to perform a given self-transformation; and (ii) obviously, such a transformation is not associated with any real change in free energy before or after the transformation (i.e., $\Delta A = 0$ in principle). These features are particularly useful for the present studies. A similar self-transformation (or zero-change) scheme has also been used before to study the convergence properties of ''single-topology'' and ''dual-topology'' approaches in calculating free energy differences in molecular systems by MD simulations.[12–14]

SYSTEM SET-UP

To investigate the influence of the extent of perturbation on the estimation of free energy differences, we used part of the *Eco*RI-binding DNA $(CGCGAATTCGCG)_2$. From the coordinates of the dodecamer DNA duplex $(CGCGAATTCGCG)_2$ in standard B-form, a part of the duplex consisting of the base sequence $(CGCGAATTC)_2$ was cut out, which was within a sphere of 17-Å radius with the nitrogen of the NH_2 group of the sixth cytosine base on the second strand at the center. This cytosine base is to be subjected to the aforementioned self-transformation. This system was then energy minimized by 100 steepest descent steps for initial relaxation. To make the system electrically neutral we placed 17 Na^+ counterions on the bisector of the O1P—P—O2P angle at a distance of 3.5 Å from the phosphorous atom of each phosphate group. The counterions were subjected to a deformable boundary force[15] to keep them from flying away. The cytosine base that was subjected to a self-transformation was replaced by a hybrid part consisting of the perturbed part in both the reactant and the product states. Different independent simulations were performed in which varying extents of the particular cytosine base were perturbed to accomplish the same self-transformation (cyt → cyt). The degree of perturbation used in achieving a transformation can be conveniently characterized by the number (n_p) of atoms that are perturbed to reach the product state. In the present work we performed calculations for four different cases in which (i) $n_p = 1$, (ii) $n_p = 3$, (iii) $n_p = 6$, and (iv) $n_p = 12$ atoms of the cytosine base were perturbed for accomplishing the cyt → cyt self-transformation. Figure 1 schematically shows the atoms that were chemically perturbed in each of these four different cases. In each of these cases, we replaced the respective cytosine base with a hybrid intermediate system consisting of the respective perturbed part in both the reactant state and the product state. Thus, in the intermediate

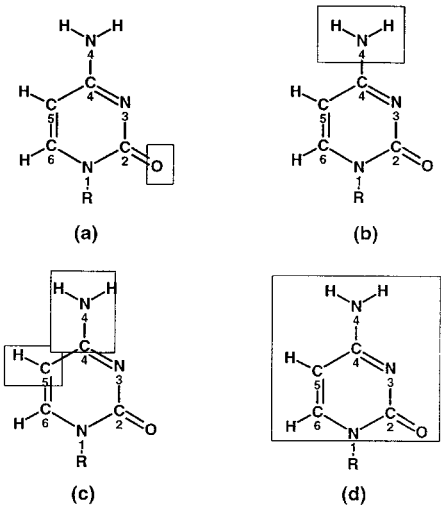

FIGURE 1. Schematic diagram showing the different atoms (boxed areas) that were perturbed for the various self-transformations used in this work. (a) $n_p = 1$; (b) $n_p = 3$; (c) $n_p = 6$; and (d) $n_p = 12$. In each case, the intermediate hybrid system contains the perturbed atoms in duplicate representing their reactant and product states.

system, the perturbed atoms exist in duplicate. In this set-up we have not considered solvent water explicitly, but have taken the solvent effect partially by performing Langevin dynamics simulation.

For comparing the magnitudes of sampling errors in the two halves of a thermodynamic cycle in the free energy perturbation calculations, we considered the case of mutation of the *Eco*RI–DNA complex as a model system. In this mutation, the NH_2 group of the adenine base at the fifth base position of the first strand is replaced by a hydrogen atom so that one hydrogen bond crucial to the DNA–protein interaction is not formed in this mutant.[16] Thus, we have prepared two systems—in one case, starting from the energy-minimized crystallographic coordinates of the *Eco*RI–DNA complex, we cut out a part that was within a sphere of 17-Å radius with the nitrogen of the NH_2 group (the mutation site) of the fifth adenine base on the first strand at the center. This adenine base was replaced by a hybrid system consisting of the NH_2 group for the reactant state and a hydrogen for the product state. Three atoms (C6, C5, and N1) nearest to the product atom were considered the colocated atoms whose charges were different in the reactant and product states. The partial atomic charges of the altered adenine with the product atom were calculated by the MOPAC 6.0[17] package by using the AM1 parameter set and ESP method. As a working approximation, we used the calculated partial atomic charge of 0.22 for the product hydrogen atom and distributed the difference in the total charge of the product and colocated atoms between their calculated charges and the corresponding charges in the CHARMM topology, equally on the calculated charges of the colocated atoms to keep the net charge the same. Thus, the partial atomic charges used for these atoms in the product state were 0.35(C6), 0.16(C5), and −0.81(N1), whereas in the reactant state these were 0.43(C6), 0.23(C5), and −0.74(N1), respectively. The partial atomic charges of the other atoms in the system were kept as usual CHARMM charges. This system, consisting of the hybrid part, was then immersed in a TIP3P[18] water sphere of 19-Å radius with the NH_2 group of the hybrid system at the center and the water molecules whose oxygen atoms were within a distance of 2.8 Å from any nonhydrogen atom of the complex were removed. To make the system electrically neutral, we placed 13 Na^+ counterions by replacing 13 water molecules whose oxygen atoms had the highest electrostatic energies. The system was then energy minimized by 500 steepest descent steps, keeping the ''reactant'' and ''product'' atoms of the hybrid part fixed to exclude the interaction between them. The energy-minimized system was then used for the MD simulation for free energy perturbation calculations. In a similar way, the DNA in the B-form corresponding to the DNA in the complex and containing the same hybrid part as in the complex was solvated in a TIP3P water sphere of 19-Å radius and 17 Na^+ counterions were included for elecroneutrality and were placed in the same way as mentioned before. The system was then energy minimized by 500 steepest descent steps, keeping the ''reactant'' and ''product'' atoms of the hybrid part fixed, and was used in the MD simulation. In a similar fashion, two additional systems were prepared corresponding to the mutant complex and free mutated DNA systems for self-transformations. The set-ups were identical, as described earlier, except for the hybrid part. In self-transformations, the hybrid part contains the NH_2 group of the respective adenine base in both the ''reactant'' as well as in the ''product'' states.

SIMULATION PROTOCOL

In cases in which different extents of perturbations were used to accomplish the same self-trans-

formation, we performed molecular dynamics simulation of the DNA by a Langevin dynamics algorithm for sampling at different intermediate states using the standard CHARMM parameter set[19] and the free energy perturbation method.[20]

The simulations were performed with a time step of 2.0 femtoseconds (fs). We used a 13.0-Å cutoff for the nonbonded interactions, and the nonbonded interaction list was updated every 10 steps. The SHAKE[21] algorithm was used for constraining all bonds involving hydrogen atoms. The system was connected to a heat bath at a temperature of 300 K. All the non-hydrogen atoms were assigned to a friction coefficient of 50 ps^{-1}. The system was first equilibrated at 300 K with $\lambda = 0.5$ for 50 ps and then was simulated in five consecutive windows at five λ values 0.1, 0.3, 0.5, 0.7, and 0.9, respectively, and in each window a 20-ps equilibration followed by a 40-ps production run was performed. We kept the bond term and the bond-angle term in the potential energy function unperturbed to maintain the structure of the perturbed part of the system for λ values close to the limiting values 0 or 1. Double-wide-sampling method was used over the full range of λ to calculate the overall free energy difference in the transformation.

For the cases of actual mutations in the *Eco*RI–DNA complex, and in free DNA in the thermodynamic cycle, we performed stochastic deformable boundary dynamic simulations in the presence of explicit water molecules. The atoms that were in the spherical shell (the buffer region) between radii 17 Å and 19 Å executed dynamics according to the Langevin dynamics algorithm, while the atoms inside the sphere of the 17-Å radius were subjected to ordinary molecular dynamics following a leap-frog algorithm. For sampling at different degrees of transformations we performed dynamics simulations at seven different values of the coupling parameter λ (0.05, 0.125, 0.25, 0.50, 0.75, 0.875, and 0.95), using SHAKE, the same nonbonded cutoff, nonbonded update frequency, and an identical protocol for the equilibration and the production runs. We performed three independent simulations on each of these systems starting with a different initial velocity distribution to get an idea of the fluctuation in the computed values.

ANALYSIS

For investigating the magnitude of sampling errors in the two halves of a perturbative cycle we used the self-transformation scheme as mentioned earlier to take advantage of the fact that, because in such cases the actual free energy difference between the two states is ideally zero, the calculated value of ΔA provides a direct measure of the error introduced by the simulation protocol. One can define the error as $\varepsilon = \Delta A_{\text{calculated}} - \Delta A_{\text{actual}}$. So, one can perform the free energy calculation on a system for a given chemical perturbation and get an estimate of the free energy difference between the two states of the system. Then, one can perform a similar free energy calculation for a self-transformation in the same system with the same reactant part, using the same simulation protocol. In such a case $\Delta A^0_{\text{actual}} = 0$, the error, ε, can then be obtained from the aforementioned relation as $\varepsilon = \Delta A^0_{\text{calculated}}$. The errors from a given simulation protocol, in general, are expected to vary for different systems. Thus, in a thermodynamic cycle, one can calculate the errors in the estimated free energy differences due to a finite simulation and the quantity $\Delta\Delta A$ can be corrected as follows:

$$\Delta\Delta A = \Delta A'_{\text{calculated}} - \Delta A''_{\text{calculated}} - (\varepsilon' - \varepsilon'') \tag{5}$$

where ε' and ε'' are the errors in the calculated values of $\Delta A'$ and $\Delta A''$, respectively, estimated by self-transformations. In the present work we were interested in verifying if $\varepsilon' - \varepsilon''$ is, in general, really insignificant compared with the computed value, $\Delta A'_{\text{calculated}} - \Delta A''_{\text{calculated}}$.

Results and Discussion

Figure 2 shows the comparison of the different ΔA vs. λ curves obtained for the four different cases of free energy calculations of self-transformation mentioned before. As in a self-transformation, the endpoints were the same, ΔA in each case should be ideally equal to zero, in principle. The calculated values of the free energy differences, ΔA, between the endpoints in the different cases studied are given in the Table I, and did not show any significant systematic trend of dependence of ΔA on the extent of perturbation in practice, as is expected due to the fact that free energy is a thermodynamic state function. However, the actual path of the cumulative ΔA vs. λ curves was clearly found to depend on the extent of perturbation (Fig. 2). The greater the extent of perturbation, the greater the maximum deviation of the curve

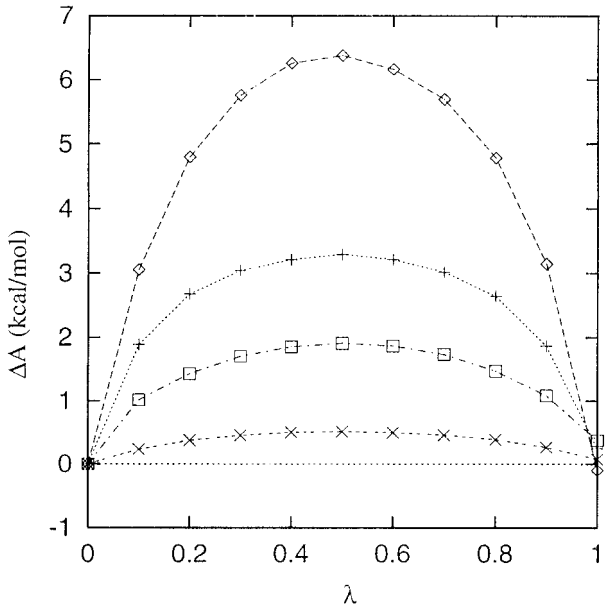

FIGURE 2. Comparison among different $\Delta A-\lambda$ curves obtained for different extents of perturbing parts ($\diamond$: $n_p = 12$; $+$: $n_p = 6$; $\square$: $n_p = 3$; $\times$: $n_p = 1$) used for accomplishing the same self-transformation in a DNA duplex system in free energy calculation.

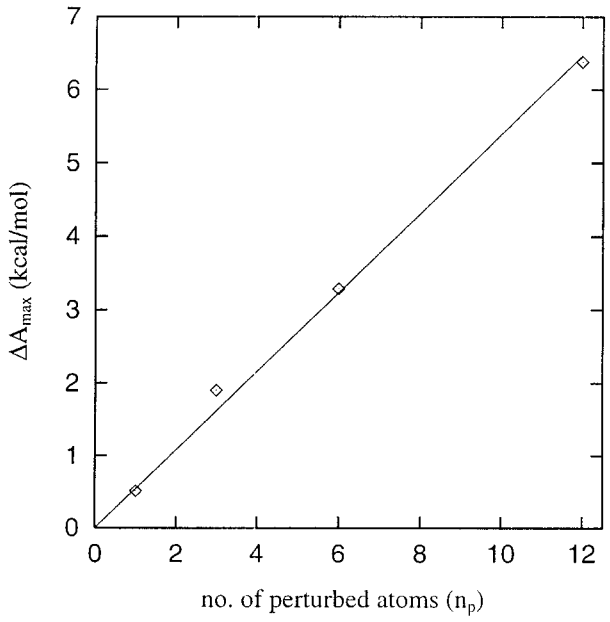

FIGURE 3. Plot of the maximum value of the accumulated free energy difference (ΔA_{max}) from the $\Delta A-\lambda$ curve for each case against the number (n_p) of perturbed atoms in the free energy calculation of self-transforming systems. The line is the linear fit of the data points showing the close linear nature of the dependence of ΔA_{max} on n_p.

TABLE I.
Comparison among Calculated ΔA Values for Different Extents of Chemical Perturbations for Same Self-Transformation and RMSD between Symmetric Points between the Two Halves of ΔA vs. λ Curve in Each Case.

No. of Perturbed Atoms (n_p)	Calculated ΔA (kcal / mol)	Normalized RMSD* in $\Delta A-\lambda$ curve
1	0.08	0.074
3	0.38	0.090
6	0.25	0.035
12	−0.09	0.012

[a] Calculated normalized RMSD between the symmetric points in each $\Delta A-\lambda$ curve.

from the zero line. In other words, the extent of perturbation actually acts as an additional degree of freedom for the system and, as a result, makes the path of the cumulative ΔA vs. λ curves dependent on the extent of perturbation used. Figure 3 shows the plot of the ΔA_{max} vs. n_p, where a fairly linear dependence is quite apparent.

It is of interest to discuss the physical basis of the shape of the cumulative ΔA vs. λ curve in the case of a self-transformation. Because, in a self-transformation there is no real transformation, the Hamiltonian of the system during the perturbative transformation is always expected to remain independent of the coupling parameter, λ, as it appears from eq. (1). It should be noted that a dual-topology implementation[20] is used in CHARMM and, in the dual-topology implementation of free energy perturbation calculations, both the reactant and the product atoms coexist separately in the molecular topology with interactions weighted according to the degree of transformation at that moment. Thus, in CHARMM, the Hamiltonians of the reactant and product parts are scaled by $(1 - \lambda)$ and λ, respectively, in the hybrid Hamiltonian to take into account the degree of transformation at that intermediate state. As a consequence, it appears that the reactant and the product parts mutually compensate the variation of the effects of each other and keep their total contribution fixed in the Hamiltonian, making the hybrid Hamiltonian independent of λ for a self-transformation case, and hence it could be expected that the ΔA vs. λ curves in all the cases should graze the $\Delta A = 0$ line. However, in reality it is not the case (see Fig. 2). The physical basis of the observed difference is simple, as discussed in what follows.

At $\lambda = 0$, the perturbed part is only in the reactant state, and at $\lambda = 1$ it is only in the prod-

uct state. At intermediate values of λ ($0 < \lambda < 1$) the perturbed part physically coexists with differentially weighted respective parts of the Hamiltonians. For the atoms in the ''reactant'' state, the Hamiltonian is scaled by a factor $(1 - \lambda)$, which implies that the force generated by each energy term in the potential energy function[14] is also scaled by a factor $(1 - \lambda)$. On the other hand, for the same values of λ, the energies and the forces acting on the atoms in the ''product'' state are obtained by scaling the Hamiltonian by a factor λ. As a consequence, two different features appear in the case of a self-transformation: (i) for different values of λ, the forces acting on an atom in the ''reactant'' (or product) state will be different and, during dynamic simulation, the associated atom will execute different λ-dependent dynamics; and (ii) the forces acting on an atom in the ''reactant'' state and the forces acting on the same atom in the ''product'' state are quite different for $0 < \lambda < 1$, except for $\lambda = 0.5$. Thus, depending on the λ value, the same atoms in the ''reactant'' state and the ''product'' state execute quite different dynamics, making the overall sampling and the energy effectively dependent on λ, and do not mutually cancel the effects of one another. For $\lambda < 0.5$, the reactant phase dominates and for $\lambda > 0.5$ the product phase dominates in the hybrid Hamiltonian, whereas at $\lambda = 0.5$, the perturbing parts in both the reactant phase and in the product phase contribute equally. Thus, the individual free energies, ΔA_is, for the different windows, are not necessarily equal to each other and close to zero; instead, the values of individual ΔA_is grow as λ approaches the value of 0.5 from either side, making the value of cumulative ΔA maximal at $\lambda = 0.5$, but keeps the total free energy change $\Delta A = \Sigma \Delta A_i = 0$ (in practice close to zero) over the full λ range. Because the ''reactant'' and ''product'' states of the perturbed part in a self-transformation are the same, approaching the point $\lambda = 0.5$ from either side ($\lambda = 0$ or $\lambda = 1$) are equivalent. As a consequence the ΔA vs. λ curve for such a transformation takes the form shown in Figure 2, and in principle should be symmetric about the point ($\lambda = 0.5$).

Clearly, with the increase in the size of the reactant and the product parts, these differences at the same λ value increase and, as a result, the deviations at the intermediate λ values are more pronounced for larger extent perturbations as compared with smaller ones. However, in practice, the ΔA vs. λ curve is not always absolutely symmetric, but slightly skewed in nature, probably suggesting a lack of adequate sampling. The RMSD between the symmetric pairs (with respect to λ) of points, on the curve normalized by the maximum value of ΔA_i, can be used as a simple measure of the degree of asymmetry or skewness between the

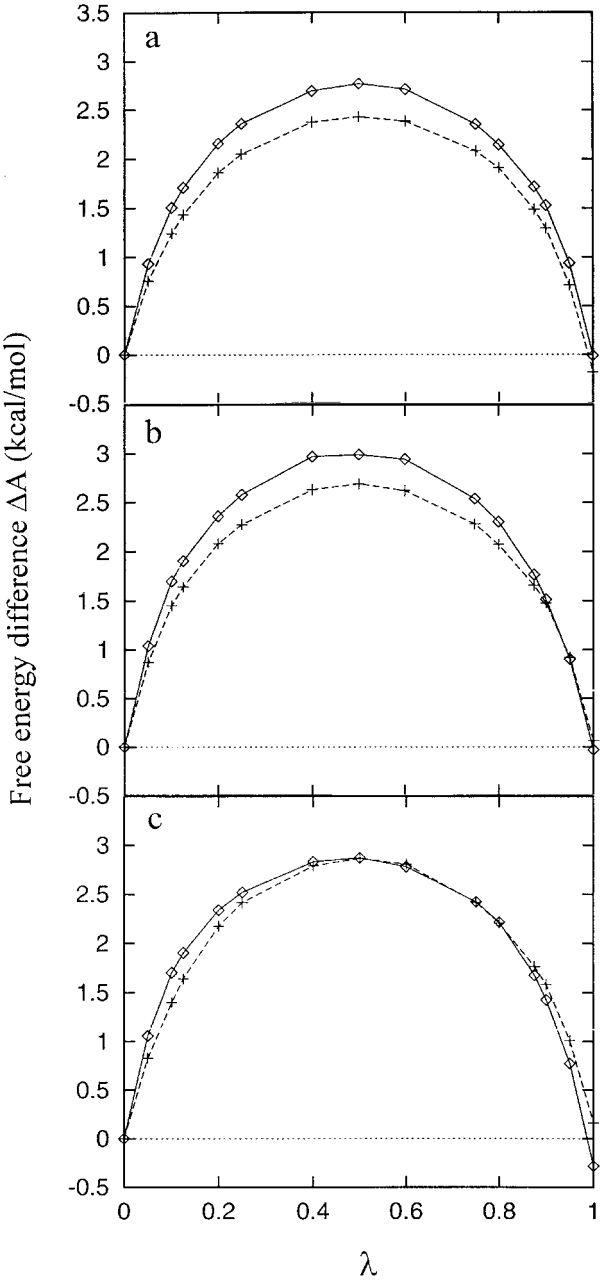

FIGURE 4. Comparison between the ΔA–λ curves for the same self-transformation performed for the free energy calculations in the free DNA (dashed line drawn through the calculated data points represented as +) and in the DNA–*Eco*RI complex system (solid line through the calculated data points represented as ◇). (a)–(c) Comparisons among three independent sets of calculations.

two halves of the curve, by the following relation:

$$\mathrm{RMSD} = \left[\left\{ \sum_{i=1}^{n} \delta\Delta A_s^2(\lambda_i, \lambda_i') \right\} / n \right]^{1/2} \qquad (6)$$

where n is the number of pairs of symmetric points on the curve considered and:

$$\delta\Delta A_s(\lambda_i, \lambda_i') = \{\Delta A(\lambda_i) - \Delta A(\lambda_i')\} / \Delta A_{\max}$$

is the difference in the ΔA values for the symmetric pair of λ values λ_i and λ_i', normalized by the $\Delta\Delta A_{\max}$ for that curve. This RMSD value may be used to assess the adequacy of sampling. Table I shows good sampling in this regard in each case. Also, on the same physical grounds, a similar behavior of the $\Delta\Delta A$ vs. λ curve is also obtained when one uses the ''single-topology'' method, but with the coexistence of the separate reactant and the product atoms as done by Pearlman.[12] Because, in such an approach the degree of transformation is taken into account by scaling each of the relevant parameters in the Hamiltonian by λ, the λ dependence of the Hamiltonian becomes different for single topology and dual topology for $0 < \lambda < 1$ and, as a consequence, the nature of the ΔA vs. λ curve for the single-topology approach looks different[12] from that obtained with the dual-topology approach as done here.

Figure 4 represents the comparison between the ΔA vs. λ curves for the self-transformations in free DNA and in DNA−*Eco*RI complex in the three independent simulations, and the Table II summarizes the relevant data. Considering the magnitude of $\Delta\Delta A$, as well as the fluctuations in $\Delta\Delta A$ between the independent simulations, it is seen that the magnitude of the differences in error $(\varepsilon' - \varepsilon'')$ is not significant. Thus, the use of thermodynamic cycles in calculating $\Delta\Delta A$ reliably from dynamic simulations in such cases appears to be quite justified in practice.

We have also calculated the ΔA values for self-transformations in the cases of the free DNA and the complex in solution, considering (i) only the first half (20-ps), (ii) the last half (20-ps), and (iii) the full (40-ps) lengths of the production run. The results are presented in the Table III. Clearly, similar fluctuations in the estimated ΔA values were found, indicating that proper equilibration was achieved in each simulation.

Summary and Conclusions

Finally, results and conclusions may be summarized as follows:

1. For a given simulation protocol, the sampling error that results due to a specific finite-length simulation does not depend significantly on

TABLE II.
Comparison between the Free Energy Data for Actual Mutation in Free DNA ($\Delta A'$) and in Complex ($\Delta A''$) and the Corresponding Self-Transformations in Free DNA (ε') and in Complex (ε'').[a]

	$\Delta A'$ (Free)	$\Delta A''$ (Complex)	$\Delta\Delta A = \Delta A'' - \Delta A'$	ε' (Free)	ε'' (Complex)	$\varepsilon'' - \varepsilon'$
1	23.8	25.2	1.4	−0.18	−0.01	0.17
2	24.3	27.1	2.8	0.07	−0.03	−0.10
3	23.2	27.2	4.0	0.16	−0.28	−0.44

[a]All the energies are expressed in kcal / mol.

TABLE III.
Comparison of ΔA Values Obtained for Different Lengths of Production Run and Also for the Same Lengths but from Different Shorter Parts of the Production Run in Cases of Self-Transformations in Free DNA and in Complex.

	Free DNA in Solution			Complex in Solution		
	First Half	Second Half	Full	First Half	Second Half	Full
Set I	−0.08	−0.28	−0.18	−0.01	−0.01	−0.01
Set II	0.65	−0.55	0.07	−0.33	0.27	−0.03
Set III	0.04	0.28	0.16	−0.09	−0.47	−0.28

the extent of perturbation used to accomplish a given chemical transformation.

2. Even in the case of a self-transformation, although there is not much difference in the free energy, ΔA, between the initial and final states, the variation in the cumulative free energy difference, ΔA, depends significantly on λ, due to the λ-dependent differential dynamics of the reactant and product parts of the system in the dual-topology approach of the free energy calculation using the chemical perturbation method. This also makes the actual path of the transformation in the $\Delta A - \lambda$ plane significantly dependent on the extent of perturbation used for achieving the chemical transformation following the same simulation protocol and the maximum deviation of ΔA from the actual value varies linearly with the size of the perturbation.
3. In a thermodynamic cycle, the error due to finite-length simulation for the free system and in complex are found to be similar to each other, justifying the use of this approach in calculating $\Delta\Delta A$. Moreover, the magnitude of error was found to be negligible compared with the fluctuations that occurred in the $\Delta\Delta A$ values calculated from independent simulations from limited sampling.

References

1. Beveridge, D.; DiCapua, F. Annu Rev Biophys Chem 1989, 18, 431.
2. Straatsma, T. P.; McCammon, J. A. Annu Rev Phys Chem 1992, 43, 407.
3. Kollman, P. A. Chem Rev 1993, 22, 184.
4. Apostolakis, J.; Muegge, I.; Ermler, U.; Fritzsel, G.; Knapp, E. W. J Am Chem Soc 1996, 118, 3743.
5. Kollman, P. A. Acc Chem Res 1996, 29, 461.
6. Gilson, M. K.; Given, J. A.; Bugh, B. L.; McCammon, J. A. Biophys J 1997, 72, 1047.
7. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
8. King, P. M. In: van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J. eds. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, Vol. 2; Escom: Leiden, 1993; p. 267.
9. van Gunsteren, W. F.; Beutler, T. C.; Fraternali, F.; King, P. M.; Mark, A. E.; Smith, P. E. In: van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., eds. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, Vol. 2; Escom: Leiden, 1993; p. 315.
10. Straatsma, T. P.; Zacharias, M.; McCammon, J. A. In: van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., eds. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications. Vol. 2. Escom: Leiden, 1993; p. 349.
11. Brooks, C. L., III. In: van Gunsteren, W. F.; Weiner, P. K., eds. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications; Escom: Amsterdam, 1989.
12. Pearlman, D. A. J Phys Chem 1994, 98, 148.
13. Pearlman, D. A.; Kollman, P. A. J Chem Phys 1991, 94, 4532.
14. Pearlman, D. A. J Chem Phys 1993, 98, 8946.
15. Brooks, C. L., III; Karplus, M. J Chem Phys 1983, 79, 6312.
16. Lesser, D. R.; Kurpiewski, M. R.; Waters, T.; Connolly, B. A.; Jen-Jacobson, L. Proc Natl Acad Sci USA 1993, 90, 7548.
17. Coolidge, M. B.; Stewart, J. J. P. MOPAC 6.0, 1990.
18. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. J Comput Chem 1983, 14, 89.
19. MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, J. J Am Chem Soc 1995, 117, 11946.
20. Fleischman, S. H.; Brooks, C. L., III. J Chem Phys 1987, 87, 3029.
21. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. J Comp Phys 1977, 23, 327.